UNIVERSITY OF PENNSYLVANIA

THE RELATIVE EFFECTS OF SIZE OF TYPE, LEADING AND CONTEXT

HOWARD T. HOVDE

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

PHILADELPHIA
1930

THE RELATIVE EFFECTS OF SIZE OF TYPE, LEADING AND CONTEXT

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF PENNSYLVANIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

HOWARD T. HOVDE

Reprints from
The Journal of Applied Psychology, Vol. 13, No. 6, and Vol. 14, No. 1

PHILADELPHIA
1930

R. UNIV.
BIBLIOTHEEK
LEIDEN

THE RELATIVE EFFECTS OF SIZE OF TYPE, LEADING AND CONTEXT

Reprinted from THE JOURNAL OF APPLIED PSYCHOLOGY
Vol. XIII, No. 6, December, 1929

THE RELATIVE EFFECTS OF SIZE OF TYPE, LEADING AND CONTEXT

HOWARD T. HOVDE

University of Pennsylvania

SCOPE OF THIS INVESTIGATION

The experiment, reported in this paper, was conducted originally in an effort to determine the relative legibility of Intertype Ideal News Face Type, set under standard newspaper publishing conditions, with changes in size and leading. The results show the relative effects of size, leading and context. The experiment is divided into two parts: Part I records the direct measurement of the reading rate of 300 adults; Part II describes the readers' preferences and opinions of the legibility of the type setting.

HISTORICAL NOTES ON LEGIBILITY

A complete review of the literature on this subject would include several hundred titles. Anyone interested in more detailed references than can be included here will find two bibliographies independently compiled by Pyke (35) and by Tinker (41), with accompanying digests of the subjects. Research in this field has been undertaken by numerous educators, ophthalmologists and psychologists. The approach to the subject has been largely psychological. Banister (3), working with test type charts of ophthalmic standards, concluded that errors often attributed to physical defects of the eye are due more likely to psychological causes and that psychological tests are a particularly desirable approach to the problem.

Two favorite methods of investigation are prevalent. The first is the test for legibility of isolated letters and numerals of which the Snellen Chart of the ophthalmologist early served as an introduction to the general problem. The second is the test for legibility of continuous reading for context with which this paper is concerned.

Under the first method, the work in ophthalmology may be mentioned. Hartridge and Owen (20) discuss the merits of the Snellen and the Green test types in ophthalmic practice. Cowan (10) calls attention

to the failure of ordinary type to satisfy the psychological requirements of a test for visual acuity and makes suggestions for changes in the character and the design of letters of equal legibility. In a second paper (11) he reviews the field of test letters and geometric designs for the determination of visual acuity.

In the field of psychology, Roethlein's comprehensive study (38) of isolated letters has received wide recognition, and more recently Tinker (42–44) has been interested in the study of isolated letters, digits and mechanical signs. Crosland and Johnson (12) have studied the range of apprehension as affected by inter-letter hair spacing and by characteristics of the individual letter. In the main these tests meet the criticism of Ewing (15), in a discussion of the value of letters and characters as visual tests, who concludes that 62 per cent of normal vision is sufficient for ordinary use and that the army accepts men with 50 per cent vision of the ophthalmic standard and with 75 per cent vision for the finest rifle work. Parsons (34) rightly contends that the characteristic group of the letters are attended to in the reading process and not the individual letters. In general, the method of investigation for the legibility of isolated letters has met with criticism by those who have studied reading habits.

Under the second method, the inquiry into reading habits permits a different approach to the subject of legibility. The reading process has been studied experimentally by Javal (25), Erdmann and Dodge (14), Huey (21–23), Quantz (36), Dearborn (13), and more recently by Judd (26), Gray (18), Schmidt (40), Buswell (7–8) and others. Cattell (9), as early as 1885, produced strong evidence that reading proceeds not by successive letters but rather by whole words, phrases and sometimes short sentences as units in the perceptual process. Reading involves a series of eye movements across the line of type. The eye pauses between movements and is fixed upon a word or group of words. During these fixations comprehension takes place, as nothing is seen while the eye is in motion. At the end of the line a regressive movement is made to the first word in the next line and the eye again continues its broken sweep across the line. Experimental studies of this nature have been largely responsible for the discard of the assumption that reading takes place by the alphabetical method; namely, that the reading process proceeds by successive letters and not by word, phrase and sentence wholes.

The reading of any consecutively written material relies upon context for the derivation of its meaning. In this usage the word "context" includes that which might be amplified into the term "contextual meaning" but which is inferred by the single word. Buswell (7) has given a clear analysis of the complex character of meaning underlying the general process of thought that is included in the context, and Huey (23) has similarly interpreted this idea.

In considering the sensory content of reading material, a variety of conflicting opinions and contradictory recommendations have arisen from experimental evidence. A few of these observations may be considered.

Size and Face of Type: Pyke (35) experimented with 8 different type faces and found rather large legibility differences in one test and small and conflicting differences in the other tests. He concluded that extremely large typographical differences must be present before it is possible to discern differences in legibility. He used the tachistoscopic method of exposing the reading material and doubts whether the differences in legibility found in the laboratory experiments hold true under ordinary reading conditions. Huey (23) concluded that the size of type is perhaps the most important single factor in legibility. Roethlein (38) stated that the legibility of a face undoubtedly increases directly with the increase in the size of a face. Freeman (16) found that 7 point type was read more rapidly than 11 point type and that the smaller type face seemed to allow somewhat greater scope of fixation, but that the number of fixations per second did not increase. However, Judd (26) found that the rate of reading and span of recognition are only slightly altered by doubling or halving the body size of a face. Parsons (34) is not opposed to condensed letters and thinks there may be an advantage in them. Huey (23) found that there is probably little to be gained by increasing the distance between letters (spacing of a word) beyond that which is used in the better printed books at the present time, although when the same number of words was printed in fewer lines of the same length and the same size of type, the material was read faster in the same proportion that the lines were decreased. Tinker and Paterson (44) obtained a speed of reading record for text which was set in roman lower case, italics and all capitals, and found the relative legibility of the type setting to be in the order named.

Leading: Noticeable differences have been found in extremely wide typographical settings in size and leading. Bentley (4) investigated Monotype News Gothic set in 12 point type and photographic reductions to correspond to 9 point and 6 point. He used as a criterion the time taken to read a given amount and concluded that 7 point leading for 12 point type, with corresponding equivalents in the smaller faces, resulted in the greatest savings, averaged at the rate of 4 minutes per hour. Typical newspaper reading conditions, however, were not present during the testing. Huey (23) found that leading was of no value when the size of type was less than 1.5 mm. for lower case nonprojectors. Parsons (34) concluded that the advantages of leading are almost negligible. Kerr (28) believed that legibility appeared to increase with leading but probably decreased as soon as one line above or below came within the small fixation area of distinct vision. Two

opposing results are to be found on the question of leading in the studies of the legibility of the telephone directory. Baird (2) concluded that within certain limits, legibility varied directly with the thickness of leading. Lyon (30) subsequently found, yet without reference to Baird's work, that there was no measurable difference between 6 point solid and 6 point type set on a 7 point base (leading 1 point), but that 1 point leading had greater aesthetic appeal and was generally preferred.

Length of Line: The width of a newspaper column today is standardized. However, earlier experiments confirm certain tendencies toward this standardization. Dearborn (13) found that 75 to 85 mm. or $3\frac{3}{4}$ inches, which is about $\frac{1}{3}$ longer than daily newspapers, combined a good many advantages with a gauge of lower case n of 1.5 mm. The shorter lines are probably better than the longer lines because words in the line below are more likely to have a sense connection while otherwise they distract attention. Ruediger (39) favored a maximum of 75 to 85 mm., no matter what the subject matter. Huey (23) preferred short rather than long lines with approximately 66 mm. as the optimum and concluded that the line-length of modern newspapers are perhaps as near the optimum (for books too) as can be determined.

Margin: Like the length of the line, margins are standardized in newspapers. Legros (29) stated that a white space should logically form the boundary between adjacent areas of print. He favored a slightly wider margin where a metal rule would otherwise be considered necessary. However, the elimination of the dividing rule is not practical in newspaper publishing, as discovered by the New York Morning Telegraph, which attempted this recently.

Paper: The Report of the Committee of Illuminating Engineering Society (37) states that paper should possess a high (diffuse) reflecting power and little gloss. Any glazing is undesirable. Glossy inks have lower diffuse reflecting powers than the mat inks, hence the latter are blacker and present a greater contrast with the paper. Lyon (30) found no measurable difference between white papers and four shades of buff colored newsprint, although white paper had the greatest aesthetic appeal.

Many typographers have judged the fitness of the printed page by its aesthetic appearance and its historical origin. Morison (31), among others summarizes as follows: (*a*) the essential form of the letters should correspond with those that have been handed down to the reader historically; (*b*) the composition of the letters should be agreeable to the words; (*c*) the lower case letters should be homogeneous in line and color; (*d*) the upper case gauge should be at most slightly less than that of the lower case ascenders. The capitals must not be conspicuous and "spot" the page; (*e*) there should be no extreme contrast of the thick and thin strokes of the letters.

The approach to the various problems have been varied and the technique is open to question. As inferred, Pyke (35), after using a tachistoscopic method, doubted whether the differences in legibility found in the laboratory experiments hold in ordinary reading. Gilliland (17) recorded the reading rate and eye movements of subjects who read two kinds of material, ranging in typesetting from 6 to 54 points and 3 to 90 points paired with a 12 point control. Results were not significant between the limits of 6 to 36 points, although the rate of reading became somewhat slower with type larger than 18 point.

Empirical judgments of some writers are based only on a criterion of aesthetic value. Historical surveys, such as Blackhurst's findings (5) and Updike's review (45), present a picture of the attempt of the printing trade to solve the problem. Manufacturers of type reflect the feeling that the most legible type is the one of which the greatest quantity is sold over the longest period of time.

Experiments such as those of Dearborn (13) and Huey (21–23) have been conducted primarily with the object of analyzing the reading process. The work of Baird (2) and Lyon (30) on the telephone directory represents tests of speed and errors in finding a name and exchange number. Gray (18) studied the performance in oral and silent reading through the analysis of eye movements on a group of subjects of various intellectual attainment. Buswell (8), with a somewhat similar method and group of subjects, investigated the span of recognition for printed matter, the rate of recognition for perception units and the regularity of perception along the line of print. Judd and Buswell (27) had subjects read passages of various degrees of difficulty and familiarity, and found significant variations of reading performance in adjustment to the change in the content of the passages.

The dangers of speed reading has been discussed by Buswell (8), Judd (26), O'Brien (32), Pyke (35), and others. Similarly, the merits of loud and silent reading have been weighted. Inasmuch as silent reading forms practically the whole of adult reading in normal life, it is preferred to oral reading for test purposes, although it presents test difficulties. Maximum and normal reading has been compared by Schmidt (40), Whipple and Curtis (46) and others. Reading at the normal rate approaches the typical reading conditions, and if controllable for test purposes has a corresponding advantage.

In conclusion to this historical review of legibility, it may be stated that many conflicting and diametrically opposing results have arisen out of experimental evidence from a variety of sources. This may be explained by the nature of the observations and the limited experimental data upon which conclusions have been drawn. To this fact may be added the great number of empirical statements that have been made by printers and others who have been interested in typography and who

have judged factors of legibility by indefinite criteria. The approach to the problem should be wholly scientific and the emphasis of the psychological aspect cannot be overlooked.

DEFINITION AND MEASUREMENT OF LEGIBILITY

The term "legibility,"[1] as used in this paper, refers to the ease and accuracy with which the type setting can be read under ordinary conditions. Legibility is a factor in reading, and reading is the process of getting meaning from the printed page. Therefore, legibility may be defined as reading with ease and accuracy, where it is understood that meaningful reading takes place. This usage differentiates the term from perceptibility or recognizability, with which the experiment was not concerned.

Legibility has to do with reading in the normal and accepted manner. Thus the reader must sit in an upright position with good illumination falling upon the reading material which is held at a normal distance for accommodation to the eyes. Conditions should be typical of the actual reading situation and should be free from mechanical impositions which might tend to distract the reader.

The measurement of legibility[2] under these conditions is the

[1] Pyke (35) observed that four times as many writers have measured legibility as have defined it. The criterion of legibility was used by 36 writers in 98 years (1827–1924), whereas only 9 writers in 44 years (1878–1921) have defined legibility. Three out of every four writers have attempted to measure something the exact nature of which they have not paused to examine.

[2] Pyke (35) states that thirty-six writers have experimented and argued with fifteen different criteria. These are: "Measurement by speed of reading (by the time threshold and the amount read), the distance threshold (direct and peripheral), 'eye span,' 'illumination threshold' focus threshold, fatigue, number of eye focuses, number of eye refixations, regularity of eye-movements, reading rhythm, 'legibility coefficient,' 'specific legibility,' size of letters, by 'judgment of the trained human eye,' and by aesthetic merit. These criteria have been used indiscriminately under a variety of conditions; namely, in silent and in loud reading, at normal and at maximum speeds, using words (isolated and grouped) and letters (isolated and grouped), with daylight, oil, and gas-light, electric light and flash-light, with continuous reading,

rate of silent,[3] mature reading. Meaningful reading must take place and be checked for accuracy. The continuous text may be measured by the amount which has been read. Inasmuch as the reading material may be set in large or small type and the leading between the lines may increase the space that the eye covers in gaining meaning, the reading area is not a criterion of legibility. The measurement can be made logically by counting the number of words which can be read in 60 seconds. This measurement assumes that the writer of the reading material has been concise and economical in the choice of words, but even if the writer has been verbose the eye must transverse the wordy phrases to gain meaning. The words, by themselves and in combination, remain the basis of the perceptual units from which meaning is derived. The normal silent reading rate should give a close approximation of the legibility of the type setting.

PRELIMINARY CONSIDERATIONS

As a preliminary approach to the problem, 106 readers were given a comparative test of reading material set in two different newspaper body faces. Reader judgments were secured, but as the test was not given under standard conditions, its usefulness served only as an introduction to the problem at hand.

For practical assistance and standardized typesetting, I am indebted to Mr. Robert McLean, Vice-President of the Philadelphia Evening Bulletin, who not only appreciated the theoretic problem but offered coöperation greater than could be expected from any other source. At the time of my first contact with Mr. McLean, a change in the body face of the news columns of the page was contemplated.

---

and brief exposures varying from one-thousandth of a second to one-tenth, and contrived with tachistoscopes of four different sorts."

[3] An experiment by Schmidt (40) reveals the physiological basis of the superiority of silent over oral reading in point of both speed and comprehension. O'Brien (32) observed: "Silent reading is as a rule a more economical and effective instrument for the assimilation of the thought of a passage than oral reading."

The immediate cause of the change was the lack of legibility in the type used. This was determined by individual complaints on the part of readers, particularly noted when the dry mat for stereotyping was introduced. The dry mat or matrix may be contrasted with the wet mat which requires heat at a relatively lower pressure to form the matrix. The dry mat on the contrary requires much greater pressure (6 tons to the square inch) to form the impression on the papier-mache which is done by rolling the mat across the type form. The introduction of new type faces might not have been necessary except for this mechanical process. In the older type face, the more delicate hair lines and serifs broke down under several dry mat impressions. Multiple impressions were necessary to supply five casting rooms and a battery of 32 presses. To illustrate the manner in which a character will break down, let us take the letter O. The top and bottom of this letter, in most designs, is a hair line. The break down of this gives the appearance of two parenthesis marks. Again, a period may be battered so that it does not print an impression. The sum total of these faulty characteristics produced illegible reading and caused complaints by the readers.

The specifications for a new type face required heavier hair lines. The practical problems of stereotyping and of high speed printing demanded this. Furthermore, a similar type design was preferred because of the reading habits of the subscribers to the paper. After considerable study, it was thought best to have a type face similar to the Roman face most often used in school books. Childhood habits formed by reading this style of type face would persist, with the adult newspaper reader. No type face which was a radical departure from this specification was considered.

One rigid specification was that no type face would be accepted which was bastard within the range of type sizes customarily set for the body of a news article. This is illustrated in the cutting of type faces. In some type faces, it is necessary to redraw the designs for extreme variations in size if certain characteristics of a line of type is to be kept. This is

usually accomplished by extending or condensing letters and by changes in the ascenders and descenders of the characters.

The selection was finally narrowed down to five type styles which were set up for printing. Dry mats were made, which in turn were cast and printed according to the conditions under which they would be used. Of these five styles, the old type of the Bulletin was one. Further study proved that, of the remainder, only two styles were outstanding, as the characteristics of the others fell within these two styles. The conservation of newspaper printing space is important, and these two remaining styles offered variations in fractions of a point; namely, $6\frac{1}{2}$ and $6\frac{3}{4}$ point.

The final selection of the Intertype Ideal News Face (24) was decided upon by the publisher after all sizes were enlarged by photo-engraving and minutely compared with one another. The rejected type style proved to be bastard in the sizes larger than $6\frac{1}{2}$ point as the whole alphabet was condensed or distorted in order that the total alphabet should not be longer than in the $6\frac{1}{2}$ point size. However, the rejection was based on the additional fact that the larger type sizes seemed to "dance" before the eye if leading were not used.

Reading material for the experiment, recorded in this paper, was carefully considered. The selection of different material for the setting of each type size was necessary in order to avoid the influence of familiarity. Outstanding news stories of the year which had been published in various papers throughout the country were considered. After these had been compiled it was found that the style of writing and interest in the material varied to such an extent that it was not usable. Aesop's Fables, essays by Emerson, Schopenhauer and others were considered and rejected because of lack of similarity to newspaper stories.

Three factors were to be considered: (*a*) the material must be by the same author in order to avoid variations in the style of writing; (*b*) reading interest in the material must be sufficiently great to stimulate the normal reading rate; (*c*) the selections must be on the same level of reading difficulty.

The experience of editors of the Evening Bulletin was helpful in this regard. Over a period of twenty-five years it had been observed that dialect stories and stories in foreign background were less popular than stories of American life. A suitable story of common interest was found in Charles A. Lindbergh's "We" which had run in serial form. From this story sixteen incidents of approximately the same level of continuous interest and reading difficulty were selected.

TEST MATERIAL AND TYPE SETTING

The selection of reading material which was believed to contain items of approximately equal interest and difficulty follows:[4]

Column 1. Lindbergh's birth and family history.
Column 2. His educational background.
Column 3. A barnstorming trip with E. G. Bahl.
Column 4. Barnstorming with Lynch in Kansas.
Column 5. Flying his own ship.
Column 6. First cross country flight.
Column 7. Landing at Meridan in "Jenny."
Column 8. Storm conditions in flying.
Column 9. A negro as a paid passenger.
Column 10. Flight from Maben to Lake Village.
Column 11. Bud Gurney's parachute incident.
Column 12. Skakopee in a cloudburst.
Column 13. Information about Army Air Service.
Column 14. Flying in Nebraska.
Column 15. Flying cross country again.
Column 16. Description of aeroplane parachute.

The above headings give a brief description of the nature of the reading material. The style of writing permitted a blending of incidents with one another in a manner that allowed reading

[4] Among those who agreed with my personal judgment of the equality of the reading material were Dr. S. W. Fernberger and Mr. R. McLean and others who read the material for this purpose. Of unsolicited opinions, many of the individuals whose test results are included in the experiment stated that they believed the material to be uniformly equal in reading interest and difficulty. This fact was subsequently substantiated by inspection of the original copy as it had been prepared for printing in the different type sizes and leadings.

with continuous interest at the beginning of every column without having completed the reading of the previous column. Paragraph indentations followed the reading context as written by the author, except in one or two instances where an especially long paragraph could be divided logically in order to equalize column paragraphs. The phrase "Check mark" described under procedure, appeared at various distances from the top of the column but within a distance that the slowest reader would find the words before time was called. Only one such instruction appeared to a column.

The material was set in full newspaper page size according to the mechanical specifications of the Evening Bulletin:

Width of type page	$17\frac{1}{8}$ inches
Depth of type page	$21\frac{3}{14}$ inches
Single column width	$2\frac{1}{12}$ inches
Columns to page	8

At the top of the page in approximately $1\frac{1}{2}$-inch depth across 8 columns were printed:

TEST FOR LEGIBILITY OF TYPE

Wait for Instructions Before Reading Any Column

At the bottom of the page in approximately 2 inch depth across 8 columns was printed:

WAIT FOR INSTRUCTIONS BEFORE ANSWERING THESE QUESTIONS

And in three divisions across the page were set:

Name.....
Check Sex: Male Female
Age.....

Do you believe that you have normal vision or corrected vision with glasses?

Yes No

If not, what is the nature of your eye defect?.....

Which of the above columns seemed easiest for you to read?
Column No. .....
Why did it appear the easiest to read?.....
.....

Each of the eight columns across the page were divided by column rules in newspaper fashion. Each newspaper column was divided into two 8-inch columns of reading matter headed

Column 1 Column 2

etc., in approximately a $\frac{3}{4}$-inch space, below which appeared reading matter, separated by a hair line rule. In this manner all reading matter was separated. Columns 1 to 8, inclusive, appeared on the upper half of the page and columns 9 to 16, inclusive, on the lower half.

The length of each line was $12\frac{1}{2}$ ems ($2\frac{1}{12}$ inches) butted next to a double bevel hair line rule with a 6 point base, leaving a margin of approximately 3 points, with the exception of the wider margin on the outside of columns 1, 8, 9 and 16. The characters adjacent to either margin were flush against the column rule with intervening letter spacings accompanied by the band space of the linotype machine.

The reading material was set in the following type sizes:

COLUMN	GROUP TEST A	GROUP TEST B AND C
1	$6\frac{3}{4}$ on 7	8 on 9
2	$6\frac{1}{2}$ on 7	6 on 7
3	$7\frac{1}{2}$ on 7	7 on 8
4	6 on 7	$6\frac{1}{2}$ on 7
5	$6\frac{3}{4}$ on 7	6 on 6
6	$6\frac{3}{4}$ on $6\frac{3}{4}$	$6\frac{3}{4}$ on $7\frac{1}{2}$
7	$6\frac{1}{2}$ on $6\frac{1}{2}$	8 on 8
8	6 on 6	$6\frac{3}{4}$ on 7
9	8 on 8	$6\frac{1}{2}$ on $6\frac{1}{2}$
10	$6\frac{3}{4}$ on $7\frac{1}{2}$	$6\frac{3}{4}$ on $6\frac{3}{4}$
11	$6\frac{1}{2}$ on $7\frac{1}{2}$	$6\frac{3}{4}$ on $7\frac{3}{4}$
12	7 on $7\frac{1}{2}$	8 on $8\frac{1}{2}$
13	7 on 8	7 on 7
14	8 on 9	$6\frac{3}{4}$ on 7
15	8 on $8\frac{1}{2}$	7 on $7\frac{1}{2}$
16	$6\frac{3}{4}$ on $7\frac{3}{4}$	$6\frac{1}{2}$ on $7\frac{1}{2}$

Explanation: $6\frac{3}{4}$ on 7, refers to $6\frac{3}{4}$ point Ideal type on a 7 point base, or leaded $\frac{1}{4}$ point; 7 on 7 refers to the size of type on the same base with no leading.

The point measurement used in this experiment is a modification of the French System which was adopted by the United States Type Founders' Association in 1886. As 996 points was the nearest metric measure of 35 centimeters, the point measurement in inches is mathematically .0138648 inches, usually spoken of as approximately $\frac{1}{72}$ of an inch. The equivalents of the type sizes used in the experiment and defined by the American Type Founders' Association are:

6 point	0.083 inch
7 point	0.09683 inch
8 point	0.1106 inch
9 point	0.1245 inch

However, the measurement used on the Mergenthaler Linotype machines which were used in setting this material in the Bulletin composing room, are as follows:

6 point	0.084 inch
6½ point	0.091 inch
6¾ point	0.0945 inch
7 point	0.098 inch
8 point	0.112 inch
9 point	0.126 inch

The designation of Group Tests A, B and C is used for convenience in reference to both test subjects and material. Each group test contains an equal number of subjects participating in the experiment. Group Test A represents the original reading material set in type sequence of random choice. In Group Test B the type setting of reading material has been reversed. This was accomplished by reversing extremes of the reading rate after the results of Group Test A had been secured. Rank order 1 of the average reading rate was set in the type setting of rank order 16, and so on until rank order 8 was changed with rank order 9. Group Test C is identical with Group Test B except for a difference in subjects and point of time in which the test was conducted. The reading material in Group Tests A, B and C is the same throughout the tests inasmuch as the order of reading for each column and the context for each column has been held constant.

From the original page form of the type setting a dry mat was made, and in turn a stereotype was cast which printed on a high speed press. The ink was Charles E. Johnson's black High Speed News Ink, and the paper grade was blue white newsprint of a 32 pound base. The same material was printed on both sides of the paper in order to produce the similar grayness of a newspaper page.

SUBJECTS FOR EXPERIMENT

Three groups of 100 individuals each took part in this experiment. The performances of these groups are designated under the headings, Group Tests A, B and C. The subjects were male[5] college students in the Sophomore and Junior years, all students or prospective students in an Advertising course at the Wharton School of Finance and Commerce, University of Pennsylvania. The students whose tests were recorded were naïve relative to typographical information. Those who were tested in the course in advertising had not at the time of the test studied typography.

By using groups of 100 individuals, it was expected that the variation of individual differences and errors (both physical and temperamental) would be cancelled by the law of average and that the result would present an accurate picture of the ordinary reader. Physiological habits in reading, such as the perceptual span, the fixation pause, eye movements, refixations, the return sweep and inner speech, were considered as constant factors that would be transmitted with equal weight to each of the columns that were read by the individual subject. Results from individuals who knew that they had defective eyesight and who lacked corrected vision by use of glasses, were discarded.

Every student had been admitted to the University by means of a psychological test or the Scholastic Aptitude Test so that normal intelligence may be presumed. While the intelligence of the group may be considered higher than the average newspaper reader, the known intelligence of this group permitted a

[5] Brooks (6) states that in silent reading sex differences are not marked.

better check on the accuracy of mature reading than would be possible if a wider range of intelligence were present. Only the results of native-born English reading students were accepted.

CONDITIONS OF TESTING

The experiment was conducted over a period of two college years. Group Test A was made in the spring months of 1928 at the close of one college year, and Group Tests B and C were made in the fall months of 1928 at the beginning of a college year. The experiment was held during a classroom hour of 50 minutes. Sixty seconds was the allotted reading time to each column with a greater rest period between readings to relieve fatigue.

The reading took place between 8 o'clock in the morning and 4 o'clock in the afternoon, with the bulk of the reading appearing nearer noonday. Natural daylight was used which was reinforced by electric ceiling lights. Walls and ceilings of the rooms were painted cream. It is believed that the illumination approximated the reading conditions in which the average newspaper is read.

Typical reading conditions were present. It was expected that each subject would read naturally and in his normal manner. Instructions were repeated three times in order that there be no misunderstanding. Personal comfort was extended to the subject in permitting any reading position that was natural. A large arm rest upon each chair permitted ease of writing. The paper was held in the manner in which the reader customarily accommodated his vision to the page.

Silent reading was followed throughout. The procedure emphasized the fact that it was not a speed test and that meaningful reading must take place. Caution was exercised throughout the experiment toward this end. The group instructions were given under the heading of Procedure incorporates the test method.

PROCEDURE

The test reading was undertaken in sections of ten to fifteen individuals. The results of each individual were combined

into Group Tests A, B and C as previously indicated. The size of the section was small enough to facilitate careful direction of each individual so that the instructions did not miscarry. None of the individuals expected a test during the hour in which it was given and great care was exerted in order to avoid the appearance of a competitive test. The following instructions were given:

"I am going to ask you to do a little reading today. I have here excerpts from Lindbergh's story 'We.' You will find that the reading is interesting, but I do not want you to consider it as a test since I am not concerned with your individual ability but with the average ability of a large group of individuals. Please follow my instructions implicitly. At the outset I am going to ask you to take alternative seats. You will find this more comfortable as you will not be crowded by your neighbor. Furthermore, I must ask you not to look at your neighbor's paper at any time during the period in which we shall read the articles which I have had specially printed. You might consciously or unconsciously enter into a competition with him. However, at the end of the period I shall give you an opportunity to make comparisons if you wish."

(At this point the test papers were held up to show the group.)

"On this page are sixteen columns, each containing a different excerpt of Lindbergh's story 'We.' Wait for instructions before reading any column. I am going to ask you to read each column normally as you would read a newspaper, but be sure you read for content.

"The reading instructions for each column will be given like this: 'Get ready to read column 1; Ready; Read!' These instructions correspond to the instructions given a track athlete, but remember you are not running a race and I want you to read *normally* and for *content*. The instructions are so given that you will have the proper attitude and a similar attitude before reading each column.

"At the word 'Read' you will start to read normally and for content. In order that I may be sure that you have read for content, I have inserted the words "(Check mark)" in parenthesis within the columns. (At this point the words (Check mark) are placed on the blackboard. These words have nothing to do with the reading matter and may appear anywhere within a sentence, paragraph or column. Whenever you see these words, put a "check mark" ($\surd$) through the words in this manner. (Here a check mark ($\surd$) was placed through the words on the blackboard.)

"After following this instruction, continue to read until I say: 'Stop.' Put a cross through the last word read. (Here a cross ($\times$) was placed on the blackboard.) Be sure you put a cross through the *last word read*,

whether this word is at the first of a sentence, the middle of a sentence, or at the end of a sentence."

"When this is done, be sure to look up and away from the paper. By no means continue to read. If you are interested in the story, I shall give you an opportunity to come back to it later. I want you to look away from the paper toward the front of the room or out of the window. I ask you to do this as relaxation. If you raise your arm once, you suffer no fatigue, but if you raise your arm ten or more times, you may notice it. Similarly you may not notice eye fatigue in reading a short time, but I want to make sure that you read each column under the same circumstances, so please look away as soon as you place a cross through the last word read. I shall then ask members of the class what they have read."

(At this point the papers were distributed to the group.)

"Please do not look at the paper that I am giving you. It is printed the same on either side of the sheet, but do not look at it or fill out anything until I give you further instructions. Remember, you have nothing to gain by trying to look at the paper ahead of instructions, so please curb your curiosity. If you do look, it is like cheating in solitaire —there is no point to it."

After the papers were distributed, the substance of the foregoing instructions was repeated and questions were asked in order to clear up any vague point. Again the instructions were repeated for the third time:

"Let me repeat briefly. You are going to read column 1 in this manner. I shall say: 'Get ready to read Column 1; Ready? Read!'—at which time you will start reading, being sure that you read normally and for content. When you see the word (Check Mark), follow the instructions (this is emphasized by pointing to the blackboard where the checkmark (√) appears through the words.) This done, continue to read until I say 'Stop.' Put a cross through the last word read. Are there any questions?

"There is one thing I might add. The time given you to read is so regulated that I do not expect the fastest reader of this group to finish the column, but should I be mistaken, will you note the fact that you have read the column in less than the allotted time. Should this be the case, of course do not read into the next column.

"Are the instructions clear?"

"Get ready to read Column 1; Ready? Read!"

The stop watch was pressed simultaneously with the word "Read." During the sixty second reading interval, I watched carefully the individuals who were reading. Distractions were noted and a caution was given at the end of the reading. In

the case of an individual who had a severe cold and who was forced to take ten or fifteen seconds out of one reading period in order to clear his head, his results were discarded.

At the end of the sixty second reading period, I was insistent that the instructions were followed faithfully. Reading stopped and the individuals all looked away from their papers.

I then asked: "Mr. Blank, where was Lindbergh's father born? Mr. So and So, give the incident described in this column." If evidence was found that the reading was faulty, the result was discarded. I then looked at each paper to see that the check mark was properly made and asked for questions on the procedure.

Before reading Column 2, the group was again cautioned that they should read normally as they would read at home without these conditions being imposed upon them.

The rest period between columns was not timed, but as the test took the full class hour, the rest period always exceeded the reading period. The remaining columns were read under the standard conditions. At the end of each reading the questions varied according to content and were also varied in the manner asked. Thus a question might be asked: "The name of Lindbergh's friend was: Smith, Lynch, Custer. Which was it?" These names had appeared in the reading, but Custer was a battlefield and Lynch was Lindbergh's friend. Again, "What type of aeroplane was used?" "Through which states did Lindbergh pass?" or "In what town did Lindbergh stop?" These specific questions were reinforced by a further statement of the incident described in each column.

At the completion of the reading of the sixteenth column, the individuals filled in the personal data relative to their age and the nature of their vision as they knew it, and made a judgment of which column seemed easiest to read and why. No clue to this latter information was given to them.

At the end, the papers were collected individually and each paper was carefully inspected to see that the instructions had been properly filled out. When known eye defects were discovered, the results of these individuals were discarded.

Furthermore, results of a Chinese and a French Canadian student were discarded because English was not their native tongue, although from casual observation their results appeared to be on a par with the others.

In obtaining the results, the following additional tests had been made, but were not used: 18 were discarded because of known eye defects; 18 were tested in the second and third experiments whose results had appeared in the first, and these results were not included because of the influence of familiarity; 23 were discarded due to errors in following instructions or reading the allotted column in less than 60 seconds; two were foreign-born individuals. Little uniformity was found among those who read the allotted material in less than 60 seconds, and the faster reading of these individuals probably resulted from a competitive spirit, skimming, lack of interest and greater interest in material, anticipating the starting signal, and similar factors which may be classed, as a whole, as errors in following instructions. No one read over half of the sixteen columns in less than the allotted time. Of this discarded group few overread more than one of the sixteen columns. The discarded papers were not tabulated.

RESULTS AND INTERPRETATION OF DATA

The 300 individual results were compiled in three groups, Group Tests A, B and C. A total of 1,318,343 words were read throughout the 4800 separate columns in the combined reading time, exclusive of rest periods, of 80 hours. While it is possible to include in the tables only the arithmetic means of the groups, anyone who may be interested may secure photostatic copies of the original data. A study of the individual records shows variations from the average but these variations are for the most part within the limitations of sampling, and the tendency of the reading rate may be well expressed by the arithmetic mean. Part I includes the results of the direct measurement of the number of words which were read; Part II describes the readers' preferences and opinions of the legibility of the type setting.

PART I

Table 1 gives a comparison of the type setting with the average number of words read in 60 seconds. The table includes Group Test A, B and C, each containing the average results of 100 readers. Besides the reading rate, the table includes the rank of the arithmetic mean, the reading order of the context,

TABLE 1

Comparison of type setting with reading rate

SIZE OF TYPE	LEADING OF TYPE	GROUP TEST A				GROUP TEST B				GROUP TEST C			
		Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preferences	Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preferences	Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preferences
6 on 6		289.22	5	8	—	258.98	12	5	—	264.89	12	5	—
	6 on 7	274.18	11	4	1	267.50	10	2	1	273.22	8	2	—
6½ on 6½		261.96	16	7	—	289.51	2	9	4	290.02	2	9	1
	6½ on 7	285.96	6	2	—	256.63	13	4	—	263.16	14	4	1
	6½ on 7½	302.47	2	11	2	251.83	14	16	—	259.38	15	16	3
6¾ on 6¾		278.47	8	6	—	275.43	7	10	—	272.67	9	10	1
	6¾ on 7	265.14	14	1	3	281.90	5	8	2	287.80	3	8	1
	6¾ on 7	273.24	12	5	—	283.31	3	14	—	277.11	7	14	1
	6¾ on 7½	277.15	9	10	1	269.17	8	6	2	281.19	4	6	8
	6¾ on 7¾	264.28	15	16	5	302.32	1	11	10	305.29	1	11	10
7 on 7		272.85	13	3	2	282.28	4	13	5	280.61	5	13	5
	7 on 7½	282.84	7	12	—	268.07	9	15	—	270.29	10	15	—
	7 on 8	290.85	4	13	11	251.36	15	3	7	263.76	13	3	11
8 on 8		306.19	1	9	30	262.16	11	7	7	266.18	11	7	4
	8 on 8½	277.02	10	15	13	277.67	6	12	32	277.12	6	12	24
	8 on 9	292.94	3	14	32	239.21	16	1	30	238.65	16	1	30

and also (this concerns Part II of the experiment) the readers' preferences as to which type face was judged easiest to read.

Table 2 contains the same data shown in table 1 with the exception that the reading order of the context has been kept in sequence and hence is a comparison of context with the reading rate.

The information contained in tables 3, 4 and 5 may be brought forward as evidence of the reliability of the arithmetic mean. Table 3 gives the means of the reading rates for each of the sixteen reading samples in Group Test A. It gives also the standard deviation of each sample and the upper and lower limits of the range in standard deviation on either side of the

TABLE 2

Comparison of context with reading rate

READING COLUMN NUMBER	GROUP TEST A				GROUP TEST B				GROUP TEST C			
	Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preference	Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preference	Av. number words read in 60 sec.	Rank order of arith. mean	Reading order of context	Readers' preferences
1	265.14	14	1	3	239.21	16	1	30	238.65	16	1	30
2	285.96	6	2	—	267.50	10	2	1	273.22	8	2	—
3	272.85	13	3	2	251.36	15	3	7	263.76	13	3	11
4	274.18	11	4	1	256.63	13	4	—	263.16	14	4	1
5	273.24	12	5	—	258.98	12	5	—	264.89	12	5	—
6	278.47	8	6	—	269.17	8	6	2	281.19	4	6	8
7	261.96	16	7	—	262.16	11	7	7	266.18	11	7	4
8	289.22	5	8	—	281.90	5	8	2	287.80	3	8	1
9	306.19	1	9	30	289.51	2	9	4	290.02	2	9	1
10	277.15	9	10	1	275.43	7	10	—	272.67	9	10	1
11	302.47	2	11	2	302.32	1	11	10	305.29	1	11	10
12	282.84	7	12	—	277.67	6	12	32	277.12	6	12	24
13	290.85	4	13	11	282.28	4	13	5	280.61	5	13	5
14	292.94	3	14	32	283.31	3	14	—	277.11	7	14	1
15	277.02	10	15	13	268.07	9	15	—	270.29	10	15	—
16	264.28	15	16	5	251.83	14	16	—	259.38	15	16	3

mean. In normal distribution 68 per cent of the cases will fall within a range of one standard deviation above or below the arithmetic mean. An examination of the seventh column of the table shows that in the sixteen samples from 62 to 70 per cent of the total cases fall within this range. The variations in the percentage of cases within the range of one standard deviation above or below the mean may be explained in terms

of the probable error of the several standard deviations. Furthermore a measure of skewness by the position of averages method showed a very small amount of assymetry. The maximal amount for any sample was −8 per cent. Hence, while this is not a complete test for normality it may be assumed

TABLE 3

Reliability of data

Group Test A

READING COLUMN	ARITH. MEAN*	S.D.	A.A. − 1 S.D.*	A.A. + 1 S.D.*	COEF. OF VAR.	PER CENT TOTAL CASES INCLUDED WITHIN RANGE OF 1 S.D. ON EITHER SIDE OF MEAN
1	265	55.44	210	320	20.9	69
2	286	51.60	234	338	18.0	61
3	273	45.26	228	318	16.6	62
4	274	50.85	223	325	18.5	71
5	273	43.28	230	316	15.9	64
6	278	50.99	227	329	18.3	67
7	262	44.03	218	306	16.8	67
8	289	57.87	231	347	20.0	70
9	306	53.94	252	360	17.6	62
10	277	56.47	221	333	20.4	67
11	302	54.71	247	357	18.1	69
12	283	59.13	224	342	20.9	68
13	291	49.19	242	340	16.9	70
14	293	51.31	242	344	17.5	68
15	277	50.73	226	328	18.3	65
16	264	49.17	215	313	18.6	69

* Decimal point omitted and fraction of 0.5 or more included as whole number.

that the distribution of the sixteen samples were approximately normal and that the variations from the average were those that could be expected in simple sampling. For the benefit of those who may be interested the coefficient of variation is given in the sixth column of the table. Tables 4 and 5 give the same information for Group Tests B and C.

Table 6 gives the correlations between the various tests by the Pearsonian coefficient formula. The correlations are all high. Group Test A represents the first test. Because of the possible influence of context, despite the fact that the material was carefully chosen, the type setting was reversed

TABLE 4

Reliability of data

Group Test B

READING COLUMN	ARITH. MEAN*	S.D.	A.A. − 1 S.D.*	A.A. + 1 S.D.*	COEF. OF VAR.	PER CENT TOTAL CASES INCLUDED WITHIN RANGE OF 1S.D. ON EITHER SIDE OF MEAN
1	239	50.25	189	289	21.0	67
2	268	46.78	221	315	17.5	65
3	251	43.12	208	294	17.1	68
4	257	45.02	212	302	17.5	71
5	259	47.77	211	307	18.4	68
6	269	54.03	215	323	20.1	67
7	262	43.46	219	305	16.6	62
8	282	57.18	225	339	20.3	66
9	290	52.58	237	343	18.1	66
10	275	64.14	211	339	23.3	66
11	302	61.36	241	363	20.3	70
12	278	61.04	217	339	21.9	64
13	282	53.62	228	336	19.0	73
14	283	60.47	223	343	21.4	70
15	268	56.02	212	324	20.9	77
16	252	47.42	205	299	18.8	74

* Decimal point omitted and fraction of 0.5 or more included as whole number.

for Group Test B and C in proportion to the extremes of the reading rate. This is equivalent to inverting the context with respect to the original type setting. In Group Test A, Column 9, set in 8 point type on an 8 point base, ranked first, and Column 7, set in $6\frac{1}{2}$ point type on a $6\frac{1}{2}$ point base, ranked

sixteenth. In Group Test B, Column 9 was set $6\frac{1}{2}$ on $6\frac{1}{2}$, and Column 7 was set 8 on 8. This change was subsequently followed by the reverse setting of Columns 2 and 14, 3 and 13, 4 and 12 until each column was completely changed to a different type size or leading. Group Test B and C represent the same material read by different individuals.

TABLE 5

Reliability of data

Group Test C

READING COLUMN	ARITH. MEAN*	S.D.	A.A. − 1S.D.*	A.A. + 1S.D.*	COEF. OF VAR.	PER CENT TOTAL CASES INCLUDED WITHIN RANGE OF 1S.D. ON EITHER SIDE OF MEAN
1	239	63.50	175	303	26.6	66
2	273	65.11	208	338	23.8	59
3	264	56.72	207	321	21.5	64
4	263	57.02	206	320	21.7	72
5	265	61.11	204	326	23.1	66
6	281	61.86	219	343	22.0	64
7	266	54.30	212	320	20.4	69
8	288	73.40	215	361	25.5	66
9	290	63.61	226	354	21.9	70
10	273	70.14	203	343	25.7	64
11	305	62.51	242	368	20.5	64
12	277	68.80	208	346	24.8	68
13	281	62.32	219	343	22.2	68
14	277	66.36	211	343	24.0	67
15	270	63.11	207	333	23.4	70
16	259	61.55	197	321	27.7	72

* Decimal point omitted and fraction of 0.5 or more included as whole number.

Group Tests B and C were performed under identical conditions except as to point of time when the tests were made. With the exception of the extremes in the reading rates Group Tests B and C vary within the limits of simple sampling.

According to the independence condition of sampling, there should be no correlation between the results of the two tests. The coefficient of correlation is, however, 95 per cent, ranked either according to the size of type or order of context. Hence we must conclude that there is some association between the two tests. There are two possible causes of relationship: (*a*) Variation of type setting, (*b*) Variations of context.

In Group Tests A and C, the context was held constant and the order of type setting was inverted. Under these conditions 74 per cent direct correlation was found between the tests.

TABLE 6

Group test correlations

Comparison of Type Setting with Reading Rate Data from table 1	
Correlations between Group Tests	A and B = −84.2 A and C = −79.8 B and C = +95.0
Comparison of Context with Reading Rate Data from table 2	
Correlations between Group Tests	A and B = +90.5 A and C = +74.0 B and C = +95.0

Correlations were computed from the mean by use of the Pearsonian coefficient formula.

Inverting the order of context and holding the order of type setting constant we find 79.8 per cent inverse correlation. Since inverting the order of context has changed a high positive correlation to an equally high negative correlation, it follows that the order of context is the dominant characteristic in determining the association between the two tests.

The conditions for Group Tests A and B, and A and C, are the same except as to point of time. We find the conclusions reached by comparing Group Tests A and C to be substantiated by a comparison of Group Tests A and B. With order of context held constant and order of type setting varied we have

90.5 per cent direct correlation. Inverting the order of context and holding the order of the type setting constant we find 84.2 per cent inverse correlation. Again it follows that the dominant characteristic is order of context.

TABLE 7

Ratio of the actual differences of the mean of each sample from the mean of all the samples to the standard error of the differences

RANK ORDER OF ARITH. MEANS	D/S.E. of Diff.		
	Group Test A	Group Test B	Group Test C
1	3.3	3.9	3.6
2	2.7	2.6	1.9
3	1.6	1.6	1.5
4	1.4	1.6	.9
5	1.0	1.5	.9
6	.7	1.0	.4
7	.3	.6	.4
8	.4	.1	.0
9	.5	.3	.0
10	.5	.3	.3
11	1.0	1.1	.8
12	1.2	1.5	.9
13	1.2	1.8	1.0
14	2.1	2.5	1.2
15	2.4	2.7	1.6
16	2.8	4.1	3.7

Computed from formula on the reliability of the difference in terms of the standard error of the difference.

It is usually customary to take D/S.E. of Diff. of 3 as indicative of complete reliability. See "Statistics in Psychology and Education" by Henry E. Garrett.

The method by which the experiment was conducted precludes the possibility of measuring the general relationship between the rate of reading and the type setting. The close correspondence between the positive correlation on the one hand and the negative correlation on the other, when the order of the context is inverted, argues fairly conclusively that no general association exists between rate of reading and the type setting.

This fact makes it necessary to turn to an investigation of the variations in the rate of reading for the individual samples.

In Group Tests A, B and C an examination of the variation in the rate of reading of each sample from the average rate of the 16 samples (table 7) shows twelve cases in which the actual differences were sufficiently in excess of the standard error of the differences to be significant. A further investigation (table 8) indicates that these significant differences are associated with the same context. Thus the context of the reading columns 9 and 11 gives reading rates outside the limits of simple sampling in all three group tests. This may be explained

TABLE 8

Significant differences in extreme samples

RANK ORDER OF MEAN	GROUP TEST A			GROUP TEST B			GROUP TEST C		
	Size of type	Context	D/S.E. of diff.	Size of type	Context	D/S.E. of diff.	Size of type	Context	D/S.E. of diff.
1	8 on 8	9	3.3	6¾ on 7¾	11	3.9	6¾ on 7¾	11	3.6
2	6½ on 7½	11	2.7	6½ on 6½	9	2.6	6½ on 6½	9	1.9
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
15	6¾ on 7¾	16	2.4	7 on 8	3	2.7	6½ on 7½	16	1.6
16	6½ on 6½	7	2.8	8 on 9	1	4.1	8 on 9	1	3.7

by the fact that these columns contained slight elements of humor and suspense, respectively, which give them exceptional interest value. The reading material in columns 1, 3, 7 and 16 was descriptive in nature, both historical and factual, but the factors which have caused these differences are not sufficiently in evidence to be determined. Likewise in the other columns no conclusion can be drawn from the relative interest or reading difficulty of the context.

Of the variables, type setting and context, context is of relatively greater importance than type setting. But as only one column is in perfect agreement among the three tests and that only as to rank order, which may be accidental, one may

inquire further as to the differences. The actual difference of the number of words read between the most read and least read columns gives this contrast:

	GROUP TEST A	GROUP TEST B	GROUP TEST C
Rank 1	306.19	302.32	305.29
Rank 16	261.96	239.21	238.65
Word difference	44.23	63.11	66.64

In reality this comparison only shows the similarity between the tests which the correlations already noted. However, it may give some indication of differences within the group. The average age of Group Test A is 20.55 years, Group Test B, 20.21 years, and Group Test C, 19.47 years. Also, Group Test A was given in the spring of the year and the other tests were given in the fall. The total number of words read throughout all columns is: Group Test A—449,476; Group Test B—431,733; Group Test C — 437,134. This factor indicates something other than the relative maturity of the subjects, as it is not coincident with it. The difference will be considered later under the general discussion of the experiment.

Possibly an interesting result, though shedding no light on the matter of size of type and leading as a factor of legibility, is found in the two columns set in $6\frac{3}{4}$ on 7. This duplication was originally inserted as a checking factor, being used in Group Test A in Column 1 and 5 where Column 1 was thought to be necessary as a "warming up" column. However, when the context was believed to be an important influence, the setting for the subsequent tests were reversed and the "warming up" column eliminated. Taken from Table Number I is the following comparison:

SIZE	GROUP TEST A	GROUP TEST B	GROUP TEST C
$6\frac{3}{4}$ on 7	265.14 (14) 1	281.90 (5) 8	287.80 (3) 8
$6\frac{3}{4}$ on 7	273.24 (12) 5	283.31 (3) 14	277.11 (7) 14
Word difference	8.10	2.41	10.69

No significance can be attached to the above information since it has been shown that the influence of context has permeated throughout the experiment. The word differences might be considered as some indication of the relative difference in context among columns 1, 5, 8 and 14. However, this information serves no purpose in determining the relative value of size and leading for the experiment as a whole.

How futile this becomes is shown in these subsequent tables:

Relationship of type size, based on rank order

SIZE OF TYPE	GROUP TEST A	GROUP TEST B	GROUP TEST C
6 on 6	5	12	12
6½ on 6½	16	2	2
6¾ on 6¾	8	7	9
7 on 7	13	5	5
8 on 8	1	11	11

It has generally been assumed by those who work with type that legibility is directly proportionate to the size and leading, within limits. In this experiment the reversal of the type setting in Group Tests B and C from Group Test A is equivalent to inverting the context with respect to the original type setting. The facts do not show a proportionate increase in legibility as the type size increases but the reading rate appears to be influenced entirely by context.

In regard to leading, this table shows a similar influence:

Relationship of leading 1 point, based on rank order

SIZE AND LEADING	GROUP TEST A	GROUP TEST B	GROUP TEST C
6 on 7	11 −	10 +	8 +
6½ on 7½	2 +	14 −	15 −
6¾ on 7¾	15 −	1 +	1 +
7 on 8	4 +	15 −	13 −
8 on 9	3 −	16 −	16 −

For convenience of interpretaton, the plus and minus signs indicate improvement and retrogression in the rank order

when 1 point leading was added. Only two out of the five cases show any improvement with leading, and if it were not for the context factor, it might be concluded that leading is of little or negative value in regard to legibility. It is evident that there is not only a lack of agreement in the influence of type sizes, but also there is a lack of agreement in the influence of leading.

In five columns, the same type size appears either on the same base or leaded from $\frac{1}{4}$ point to 1 point. The effect of leading appears in tabular form below:

Relationship of leading same type face, based on rank order

SIZE AND LEADING	GROUP TEST A	GROUP TEST B	GROUP TEST C
$6\frac{3}{4}$ on $6\frac{3}{4}$	8	7	9
$6\frac{3}{4}$ on 7	14 −	5 +	3 +
$6\frac{3}{4}$ on 7	12 −	3 +	7 +
$6\frac{3}{4}$ on $7\frac{1}{2}$	9 −	8 −	4 +
$6\frac{3}{4}$ on $7\frac{3}{4}$	15 −	1 +	1 +

The few cases of the same single type size with leading do not warrant a separate statistical correlation as the few items will result in a large probable error. With one exception the effect of leading has inverted the order of the results of Group Test A, when the context was changed in Group Test B and C. If a similar comparison is taken from the entire experiment, the effect of leading as a whole over 16 columns of all three tests, 13 examples are found where leading has aided legibility and 20 where it has retarded the reading rate. These figures are cited only generally as no conclusions can be drawn from them since we have previously demonstrated the dominant influence of context.

CONTINUATION

For convenience in publication, it has been necessary to divide the paper at this point. For the purpose of continuity, it may be noted that Part II deals with the readers' preferences and opinions of legibility in contrast with the criterion of the reading rate. The discussion of both Parts I and II along with the general conclusions and bibliography must be left to the next issue of the Journal.

Reprinted from The Journal of Applied Psychology
Vol. XIX, No. 1, February, 1930

THE RELATIVE EFFECTS OF SIZE OF TYPE, LEADING AND CONTEXT

HOWARD T. HOVDE

University of Pennsylvania

For convenience of publication, an arbitrary division of this paper was made in the last issue of the Journal. The preceding article outlined the scope of the investigation, reviewed the historical aspects of the problem of legibility, stated the procedure of the investigation and gave the results of Part I. Eight tables and several abstracts from these tables have been published. In this article it has been necessary to refer to both the discussion and tables in Part I where the measurement of legibility was taken as the rate of silent, mature reading. The present article gives the readers' preferences and opinions of legibility in contrast with the criterion of the reading rate and concludes the discussion of both parts of the investigation.

PART II

The foregoing information has been based upon the actual number of words read. This fact may be contrasted with the readers' preferences and opinions of legibility. See tables 1 and 2. The literal question was: Which of the above columns seemed easiest to read? If a division of the rank order of number of words read is divided into half, it is found, in Group Test A, that 75 prefer the first half of the columns in which the most read words were actually read, and 25 of the last half. In Group Test B, the ratio is 55 to 45, and in Group Test C, it is 50 to 50. These figures are merely compared for the group. The preferences of the subjects seldom agreed with their best reading rates. If the ratio of the readers' preferences is compared with actual words read, divided into the first five, mid-

dle six, and last five columns, the ratio is: Group Test A: 75, 14,11; Group Test B: 21, 35, 44; Group Test C: 25, 26, 49. This fact substantiates the reader's unreliability of judgment when compared with the reading rate.

If ten or more readers' preferences are recorded, it is found in Group Test A that 32 prefer 8 point type leaded 1 point; 30, 8 point solid; 13, 8 point leaded $\frac{1}{2}$ point, and 11, 7 point leaded 1 point. In Group Test B, 32 prefer 8 point leaded $\frac{1}{2}$; 30, 8 point leaded 1 point, and 10, $6\frac{3}{4}$ points solid. In Group Test C, 30 prefer 8 point leaded 1 point; 24, 8 point leaded $\frac{1}{2}$ point; 11, 7 point leaded 1 point, and 10, $6\frac{3}{4}$ point leaded 1 point. The tendency of the readers is to accept the larger type faces and leading as easiest to read, though this fact does not agree with the actual number of words read. Undoubtedly in the discrimination of small differences, the readers' opinions of the easiest reading was based upon the most space which the eye had covered although larger type and leading did not permit the inclusion of as much reading material despite the fact that the eye covered more space. Undoubtedly the columns in which the most words were read on an average, show less readers' choices than they actually deserve because the readers thought they did not cover as much ground.

The reasons for the readers' preferences were varied. In general they may be classified according to factors of size, leading and interest, and miscellaneous factors which could not be divided into a more definite category.

Size was described in varying terms such as larger, largest, larger and blacker, large and uniform, larger and stronger, etc.

Leading was described by the reader as space between lines, not crowded, spaced better and not so black, greater open space, etc.

Interest was described in terms of reading material as interesting, lighter reading material, writing flowed together, interest in material read, etc.

Miscellaneous factors were suggested as follows: Felt more comfortable while reading it; Less glaring; Clear (with no other description); Could read with less trouble; Looked like other

newspaper type; Words easiest to recognize; Understood perfectly what was read; Less confusion of print; Unable to give reasons; Easiest to read; Distinct; More legible; Less strain; Seemed less effort; Words not "jumbled;" Followed out line better; Words stand out more; More read with less

TABLE 9

Reasons given for preference

READING COLUMN NUMBER	GROUP TEST A				GROUP TEST B				GROUP TEST C			
	Size	Leading	Interest	Miscellaneous	Size	Leading	Interest	Miscellaneous	Size	Leading	Interest	Miscellaneous
1	2	—	1	—	19½	6	1½	3	11	14	1	4
2	—	—	—	—	—	—	—	1	—	—	—	—
3	1	1	—	—	2½	2½	½	1½	3	3	1	4
4	—	—	½	½	—	—	—	—	—	1	—	—
5	—	—	—	—	—	—	—	—	—	—	—	—
6	—	—	—	—	1	½	½	—	2½	½	2	3
7	—	—	—	—	4	1	1	1	2	½	½	1
8	—	—	—	—	—	—	1	1	1	—	—	—
9	13½	6	5½	4	1	—	3	—	—	—	1	—
10	½	—	½	—	—	—	—	—	—	—	1	—
11	½	½	½	½	3	1	4	2	2	—	4	4
12	—	—	—	—	15	5	6	6	8	6	5	5
13	3½	3½	—	5	4	—	—	1	2	2	1	—
14	7½	13½	2	9	—	—	—	—	—	—	—	1
15	7	4	—	2	—	—	—	—	—	—	—	—
16	½	1½	1	2	—	—	—	—	—	1	2	—
Total	36	30	11	23	50	16	17½	16½	31½	28	18½	22
	100				100				100			

effort; Very plain; Minimized eye movements; Less time; Got further; Read at beginning; Easiest to eyes; It seemed best type suited to my eyes; Eyes not too tired; Less strain; Because I like that type of print, etc.

Table 9 is a complete analysis of the readers' preferences.

Whenever dual reasons were given for a preference, the reason was scored as a fraction. Compilation was made with the column number as a unit. The actual size or leading of the type may be found by referring to the column number in table 1.

The bulk of the reasons given by the readers led to direct classification. Most readers were influenced by size of type and secondly by leading. However, interest in the context which actually dominates the amount read has been relegated to a minor position in the readers' opinions. Again this shows a tendency, developed on the part of most individuals, to believe in the judgment of space as a whole, rather than in a quantitative judgment of the amount of reading within the space. The presence of a large miscellaneous class may tend to vitiate the results if this class overlaps the others. However, it cannot be more definitely distributed as it constitutes perceptual factors indefinitely analyzed by the readers.

DISCUSSION

Two deductions can be made from this experiment: (*a*) Context is a more important factor in determining legibility than the sensory content of the type setting when the reading rate is used as a criterion; (*b*) The sensory content is the basis of the readers' preferences and opinions of legibility.

In the first part of this experiment the results are opposed to what is customarily accepted in typography—namely, that legibility is increased, within limits, when the size of type and leading are increased. No deduction can be made from the value of changes in the sensory content as long as context has so great an influence. Perhaps future work in this field can establish a value proportionate to the influence of context through a weighted statistical treatment and in this manner arrive at the relative value of the sensory content. Nevertheless, these observations are presented for their present value. Any treatment of reading material attempting to determine the value of the sensory factors will be tainted by context in the same manner in which this problem has been. Many of the past experiments in legibility have been influenced by context when

meaningful reading has been employed. To attempt to avoid this influence by the use of nonsense syllables or isolated letters would nullify the object of the experiment and would test some other factor than legibility as it has been defined. Identical material can never be used because familiarity has a great influence on the rate of reading.

It is apparent from experiments in this field that many variables have entered into problems on legibility and controls have been difficult to establish. The experiment, reported in this paper, has experienced similar difficulties, but the treatment of the results has been by the method of the arithmetic means of groups of 100 subjects, which has reduced the probable error. Few former experiments have used more than eight or ten subjects and seldom over fifty subjects.

It is interesting to note the high correlation between the group tests. The correlation between the Group Tests B and C is 95 per cent direct correlation.[1] The coefficient of correlation of both of these tests with Group Test A is somewhat lower, although sufficiently high to consider the results reliable. It will be noted that the average reading rate is highest in Group Test A. An explanation for this may be found in the results of Packer and Anderson (33) in their work on the effect of summer vacation upon the rate of reading. Their observations were confined exclusively to children from the first to the fifth grades and their results show a corresponding decrease in the reading rate in tests conducted in September as against the tests conducted in May. In this experiment Group Tests B and C were made in the fall of the year and Group A in the spring. Inasmuch as a college recess period intervened between the tests, it is probable that the summer vacation period affects college students in a manner similar to that of grade school children. This observation is further substantiated by a rise in the total number of words read throughout the sixteen columns of Group Test C when compared with Group Test B. Group Test C followed Group Test B by about one month in the college year.

[1] Cf. Table 6 and discussion on page 624, this Journal, 1929, Vol. XIII, No. 6.

The difference in total reading rates is in evidence despite a slight drop in the average maturity of the readers' ages in Group Test C, and may be explained by improvements in the reading faculty of the students as they progress during the college year.

The uncontrollable variable of this experiment has been the reading material. A careful choice of this material was made originally, and previously judged as nearly equal in interest and reading difficulty as it was possible to secure. The context was simple and the style of writing was the same throughout. It is well known that simple material will be read more rapidly than difficult material, but it was thought that essentially similar material would have a constant reading rate. Judd and Buswell (27), working with university students who read long passages of different kinds of subject matter including poems, prose, text-books and foreign language reading, found that different reading attitudes were induced by different material. While it would be expected that difficult reading material would decidedly decrease the reading rate, Huey (23) had found that when individuals read silently by their own method they fell into a natural rate which was nearly uniform for page after page. Huey's statement must be qualified inasmuch as it now appears that the reading rate varies directly with context in minute changes of reading interest and difficulty and continues to vary notwithstanding the changes in the sensory content. It is apparent that if there is no uniform mechanical procedure in reading difficult material this same lack of uniformity will be relatively true in reading the most simple material despite the fact that the simplicity of the context is essentially similar.

Another factor underlying the reading attitude may be found in the interpretation of skimming by Whipple and Curtis (46). They defined different modes of rapid reading and skimming as follows: (*a*) Fast reading; (*b*) "Trailing" which is "perceiving without apperceiving," catching impressions which may or may not be meaningful for context—probably not in most cases; (*c*) "Covering," like trailing, except that it may be either up or down or right and left, while trailing is the latter; (*d*) Omitting logically; (*e*) Omitting arbitrarily. From the average time of all

observers they found that the order of speed of five methods of reading are: Skimming, maximal silent, normal silent, maximal aloud and normal aloud. Knowledge that reproduction is to be demanded slows the rate of reading of all subjects by all methods. Inasmuch as reading instructions may stimulate a different reading rate, it is probable that the interest content in a reading passage likewise may cause a difference.

A few previous experiments falling into this error may be mentioned. Gilliland (17) experimented on the effect on reading of changes in size of type, using both adults and children as subjects. Two different sets of paragraphs with relatively large changes in type size were used, but the results were influenced very probably more by context than by changes in size, particularly in the sizes under 12 point. Similarly the conclusions reached by Griffing and Franz (19) are doubtful. They had observers read two passages of the Bible containing 622 words, set in Roman type (not otherwise described) in Pica size (about 12 point) and Pearl (about 5 point). Their conclusion in favor of the larger type size possibly has been influenced by context more than by sensory content. The studies by Huey (23) have been largely confined to the reading process but his recommendations for type size may be influenced by context rather than type size and leading since he has used the reading rate as a criterion. Experimental evidence which has resulted primarily from the study of reading habits tends to confirm the importance of context as a more important factor of legibility than the type setting.[2]

The second part of this experiment has to do with an empirical judgment of the readers' opinions of legibility. The largest sensory variation, i.e., that between type set in the smallest size and largest size with leading, was 3 points, or approximately $\frac{1}{2}\frac{1}{4}$ of an inch; the smallest difference is $\frac{1}{4}$ point, or about $\frac{1}{2}\frac{1}{8}\frac{1}{8}$ of an inch. These variations are well within the limits in which the eye can detect differences. When the subject was asked "which column seemed easiest to read," it appears that his judg-

[2] See Buswell (8), Judd (26) and O'Brien (32).

ment was based on a space relationship, carrying a picture in his mind of the space which his eye had covered as the nearest identity to the amount which was read and giving as the reason for his judgment the size of type and leading. This space concept makes many readers turn away from closely printed matter to more open printing, believing it is easier to read. This space illusion may be a corollary of the distance illusion of converging lines. Just as the architect places the columns of a cathedral nave closer together near the altar in order to give the impression of distance, the printer varies size of type and leading in order to deceive the reader in the amount of space which the eye covers.

Large type, leading, or both, at the beginning of a printed column or page trick the eye into reading. After interest in the subject matter has been established, it is reasonable to assume that the eye can accommodate itself, not to a decreased legibility because of reduction of size or leading, but to an increase of legibility as measured by the reading rate. Smaller type faces without leading probably increase the working efficiency of the eye due to the conservation of eye movements. Fatigue, particularly noticed in reading extremely small type, possibly is caused by the confined reading area. Attention is more closely fixed upon smaller sensory units which do not allow as great an interval of relaxation as do larger sensory units. But the rapid adjustment of the eye permits reading in both small and large type with seemingly equal facility. The greater amount of reading work done within the area covered by smaller type setting is seldom recognized by the reader in forming a judgment of legibility. The inherent attention value of context dominates the sensory content of reading material so that within the normal reading rate, the reader's interest in the context speeds up the reading process and the reader reads most of that which interests him most.

CONCLUSIONS

1. From the results of this experiment it may be concluded that the context of the reading material is a more important

factor in its effect on the reading rate, and consequently upon legibility, than the sensory material. Changes in size of type and leading are not as dominant an influence as changes in the context of the reading material. Because of the dominance of context, no conclusion can be drawn from the relative effects of changes in size of type and leading. Experimental evidence by those who have studied reading habits corroborates this conclusion.

2. The readers' preferences and opinions of legibility differ from the amount of reading as measured by the reading rate. The readers' preferences for larger type and leading probably are due to an impression that more material has been read because more space has been covered in the reading, although the reasons for the preferences are given in terms of sensory content. Reasons for the choice of larger sensory units in some instances are confused with context, but opinions of legibility are stated largely in terms of sensory content. The results of this experiment do not warrant specific recommendations for the use of a particular type size or leading for newspaper columns within the limits of the type size and leading that have been tested.

REFERENCES

(1) Atkins, E. W.: The efficiency of the eye under different intensities of illumination. Jour. of Comp. Psychol., 1927, Vol. 1, 1–37.

(2) Baird, J. W.: The legibility of a telephone directory. Jour. Appl. Psychol., 1917, Vol. 1, 30–37.

(3) Banister, H.: Block capitals as a test of visual acuity. Brit. Jour. Ophthal., 1927, Vol. 11, 49–61.

(4) Bentley, M.: Leading and legibility. Psychol. Monogs., 1921, Vol. 30, No. 136, 48–61.

(5) Blackhurst, J. H.: Investigations in the Hygiene of Reading. Warwick and York, 1927, Baltimore.

(6) Brooks, F. D.: The Applied Psychology of Reading. Appleton & Co., 1926, New York.

(7) Buswell, G. T.: An experimental study of the eye-voice span in reading. Suppl. Educat. Monogs., 1920, No. 17.

(8) Buswell, G. T.: Fundamental reading habits; a study of their development. Suppl. Educat. Monogs., 1922, No. 21.

(9) Cattell, J. M.: Inertia of eye and brain. Brain, 1885, Vol. 8, 295–312.

(10) Cowan, A.: Test letters which comply with the physiological requirements of a visual test object. Am. Jour. Ophth., 1928, Vol. 11, 625–628.

(11) Cowan, A.: Test cards for determination of the visual acuity. A review. Arch. Ophth., 1928, Vol. 57, 283–297.

(12) Crosland, H. R., and Johnson, G.: The range of apprehension as affected by inter-letter hair-spacing and by characteristics of individual letters. Jour. Appl. Psychol., 1928, Vol. 12, 82–124.

(13) Dearborn, W. F.: The psychology of reading. Columbia Univ. Studies, Arch. of Phil., Psychol. and Scien. Methods, 1906, Vol. 14, No. 1.

(14) Erdmann, B., and Dodge, R.: Psychologische Untersuchungen über das lesen auf experimenteller Grundlage, Halle, 1898.

(15) Ewing, A. E.: The value of letters and characters as visual tests. An International Congress of Ophthalmology, Washington, D. C., 1922, 604–606.

(16) Freeman, F. N.: Experimental Education, 1916, chap. 3, experiments 9–11. Houghton Mifflin Co., Boston.

(17) Gilliland, A. R.: The effect on reading of changes in the size of type. El. Sch. Jour., 1923, Vol. 24, 138–146.

(18) Gray, C. T.: Types of reading ability as exhibited through tests and laboratory experiments. Suppl. Educat. Monogs., 1917, No. 5.

(19) Griffing, H., and Franz, S. I.: On the conditions of fatigue in reading. Psychol. Rev., 1896, Vol. 3, 513–530.

(20) Hartridge, H., and Owen, H. B.: Test Types. Brit. J. Ophthal., 1922, Vol. 6, 543–549.

(21) Huey, E. B.: Preliminary experiments in the physiology and psychology of reading. Amer. Jour. Psychol., 1898, Vol. 9, 575–586.

(22) Huey, E. B.: On the psychology and physiology of reading. Amer. Jour. Psychol., 1900, Vol. 11, 283–302; 1901, Vol. 12, 292–313.

(23) Huey, E. B.: The Psychology and Pedagogy of Reading. 1908, Macmillan Co., New York.

(24) Intertype: Book of Intertype Faces. Intertype Corporation, 1928, New York.

(25) Javal, E.: L'evolution de la typographie consideree dans ses rapports avec l'hygiene de la vue. Rev. Scie., Troisieme Serie, 1881, Vol. 1, 802–813.

(26) Judd, C. H., and Others: Reading: its nature and development. Suppl. Educat. Monogs., 1918, No. 10.

(27) Judd, C. H., and Buswell, G. T.: Silent reading, a study of the various types. Suppl. Educat. Monogs., 1922, No. 23.

(28) Kerr, J.: School Vision and the Myopic Scholar. 1925, Chap. 5, London.

(29) Legros, L. A.: Note on the legibility of printed matter. Committee on type faces, H. M. Stationary Office, 1922, London.
(30) Lyon, O. C.: The Telephone Directory. Bell Tel. Quart., 1924, Vol. 3, 175–185.
(31) Morison, S.: Towards an Ideal Type. The Fleuron, 1924, Vol. 2, 27.
(32) O'Brien, J. A.: Reading, its Psychology and Pedagogy. 1926, Century Co., New York.
(33) Packer, P. C., and Anderson, H. W.: Loss in reading ability during summer vacation. Midland Schools, 1915, Vol. 30, 81–82.
(34) Parsons, J. H.: A review of report of committee appointed to select best faces of type and display for government printing and of note on legibility of printed matter by L. A. Legros. Brit. Jour. Ophthal., 1922, Vol. 6, 475.
(35) Pyke, R. L.: Report on the legibility of print. 1926, H. M. Stationery Office, London.
(36) Quantz, J. O.: Problems in the psychology of reading. Psychol. Rev., Monog. Suppl., 1897, Vol. 2, no. 1.
(37) Report of the committee on glare of the illuminating engineering society; New York, 1915. Diffusing Media III, Papers and Ink, Tr. Ill. Eng. Soc., Vol. 10, 379.
(38) Roethlein, B. E.: The relative legibility of different faces of printing types. Amer. Jour. Psychol., 1912, Vol. 23, 1–36.
(39) Ruediger, W. C.: The field of distinct vision. Col. Univ. Arch. of Psychol., 1907, Reprints, No. 5, pp. 67.
(40) Schmidt, W. A.: An experimental study in the psychology of reading. Suppl. Educ. Monog., 1917, No. 2.
(41) Tinker, M. A.: Legibility and eye movement in reading. Psychol. Bull., 1927, Vol. 24, 621–639.
(42) Tinker, M. A.: Numerals versus words for efficiency in reading. Jour. Appl. Psychol., 1928, Vol. 12, 190–199.
(43) Tinker, M. A.: The relative legibility of the letters, the digits, and of certain mathematical signs. Jour. Gen. Psychol., 1928, Vols. 3 and 4, 472–494.
(44) Tinker, M. A., and Paterson, D. G.: Influence of type form on speed of reading. Jour. Appl. Psychol., 1928, Vol. 12, 359–368.
II. Size of type. Ibid., 1929, Vol. 13, 120–130.
III. Length of line. Ibid., 1929, Vol. 13, 205–219.
(45) Updike, D. B.: Printing Types. 2 vols. Harvard University Press, 1922, Cambridge.
(46) Whipple, G. M., and Curtis, J. N.: Preliminary investigation of skimming in reading. Jour. Educat. Psychol., 1917, Vol. 8, 333–349.